An Evaluation of Selected Erosional Factors on Grand Canyon Archeological Sites

The purpose of this paper is to assess the relationship of set of variables to the degree of observed erosion of the 151 Grand Canyon archaeological sites included in the current Reclamation treatment plan in compliance with section 106 of the National Historic Preservation Act of 1966 (as amended).

Damp, Pederson and O'Brien (2007) have evaluated these sites and ranked the erosional vulnerability from 1 (least) to 5 (greatest) based on a combination of geomorphological and archeological criteria. Site types unlikely to be affected by flow regimens (i.e., isolated pots, caches, rock art, and historic inscriptions) were removed from the analyses resulting in a reduced site count of 133.

The statistical method employed is discriminant function analysis (Klecka 1980) using five continuous independent variables with the site condition ranking treated as dependent categorical variables. That is, we consider the relative affects of environmental variables on categories of erosional vulnerability. Data were analyzed with the discriminant function option of the Statistical Package for the Social Sciences (SPSS), version 15.0 for Windows.

The Sites

Site types include pueblos (2), small structures (21), ephemeral structures (7), storage structures (1), enigmatic features (2), sherd scatters (1), lithic scatters (1), multiple class artifact scatters (3), isolated thermal features (4), roaster complexes (47), prehistoric camps (34), trails (1), historic scatters (1) and historic camps (6). These are located between river mile 1 and 259. Sites are distributed tri-modally with major site density peaks at 75 and 200 miles and a minor peak at 125 miles.

Independent Variables

Independent variables included in the analysis are 1) river mile, 2) elevation ASL, 3) distance from river, 4) slope and 5) HECRAS threshold. The first four variables were taken from the Grand Canyon Microsoft ACCESS archeological database. The fifth variable was generated by GCMRC and represents inundation flow models. Discriminant function analysis assumes a multivariate normal distribution of the independent variables. Histograms of the raw variables indicated marked skewing and multi-modality hence it was necessary to perform normalizing transformations prior to analysis. This was achieved using two types of transforms. Simple power transforms were applied to river mile and elevation as follows where:

$$T(X) = X^n$$

River Mile

$T(RM) = RM^{4.3}$

Elevation

$T(E) = E^{3.3}$

Box-Cox power transforms were applied to distance from river, slope and HECRAS threshold as follows where:

$T(X) = (X^{\lambda} - 1) / \lambda$

Distance from River

$T(DR) = (DR^{0.42} - 1) / 0.42$

Slope

$T(S) = (S^{-0.5} - 1) / -0.5$

HECRAS

$T(H) = (H^{2.5} - 1) / 2.5$

The transformations to normal distributions were not ideal due, in large part, to the multi-modal nature of the variables, but they provided a significant improvement over the raw data. Both types of transforms were accomplished by repeated changes in values of the exponents followed by a display of the resultant histograms until reasonably normal distributions were observed.

Dependant Categorical Variables

The five erosional categories developed by Damp, Pederson and O'Brien (2007) were, for present purposes, collapsed into two categories, combining levels 1, 2 and 3 into category 1 and levels 4 and 5 into category 2. This approach was taken because of the relatively small sample size (133) of historic properties to be analyzed and the resultant difficulty of meaningfully discriminating among 5 distinct categories.

Analytic Results

SPSS discriminant was run with the following options: Calculate independent variable means, Box's M, Fisher's linear discriminant coefficients and unstandardized coefficients (Landau and Everitt 2004). Analysis was performed considering all independent variables together rather than a stepwise approach. Prior probabilities were based on group size and classification was based on the entire data set as well as a cross-validation technique referred to as *leave one out.*

The means and standard deviations of the transformed variable are shown below.

Group Statistics

NewDependent		Mean	Std. Deviation
1	N_RiverMile	183273423.41	929533722.089
	N_Elevation	188126241.25	1146412648.497
	N_DistRiver	11.24	4.133
	N_Slope	.91	.640
	N_Hlevel	60110585.48	1044349627.491
2	N_RiverMile	135367655.62	366162548.989
	N_Elevation	-282312302.72	1159358956.817
	N_DistRiver	11.90	3.177
	N_Slope	1.24	.511
	N_Hlevel	96270266.21	1022285208.854
Total	N_RiverMile	172827804.87	838475959.498
	N_Elevation	85549415.87	1161317123.492
	N_DistRiver	11.38	3.942
	N_Slope	.98	.627
	N_Hlevel	67995027.14	1035835706.059

Box's Test of Equality of Covariance Matrices
Log Determinants

NewDependent	Rank	Log Determinant
1	5	126.441
2	5	123.519
Pooled within-groups	5	126.078

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results

Box's M		34.328
F	Approx.	2.132
	df1	15
	df2	10578.663
	Sig.	.007

Tests null hypothesis of equal population covariance matrices.

Box's M has a significance value of .007 indicating that the data differ significantly from multivariate normal. Thus, even though the individual independent variables were normalized, a multivariate normal distribution was not achieved.

Summary of Canonical Discriminant Functions
Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.082(a)	100.0	100.0	.275

a First 1 canonical discriminant functions were used in the analysis.

The eigenvalue of the first discriminant function is only .082. Such a low value will expectably lead to relatively weak discriminating power. Wilk's Lambda leads to a similar conclusion. Wilk's Lambda is the ratio of within-groups sum of squares to the total sum of squares. This is the proportion of total variance in the discriminant scores *not* explained by differences among groups. A lambda of 1.00 occurs when observed group means are equal, while a small lambda occurs when within-groups variability is small compared to the total variability. A small lambda indicates that group means appear to differ. In the case at hand a Wilk's Lambda of .924 suggests that the two group means are nearly identical.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.924	10.143	5	.071

These weaknesses notwithstanding, the analysis does allow for the evaluation of the comparative discriminating power among the independent variables through the interpretation of the standardized canonical discriminant function coefficients. As shown below, the normalized variables elevation (-.592), slope (.774) and distance from river (.177) have the greatest power (the sign of a coefficient, positive or negative, is not relevant to discriminating power). River mile and HECRAS threshold have negligible values.

Standardized Canonical Discriminant Function Coefficients

	Function
	1
N_RiverMile	-.089
N_Elevation	-.592
N_DistRiver	.177
N_Slope	.774
N_Hlevel	-.014

It is interesting to note that the multivariate centroids of the two groups are fairly well separated. This would indicate that the relatively poor discrimination performance is due largely to the high degree of variability in the independent variables.

Functions at Group Centroids

NewDependent	Function 1
1	-.150
2	.539

Unstandardized canonical discriminant functions evaluated at group means

Finally, one test used to asses the credibility of a discriminant solution is to reclassify the original cases using the derived coefficients. A total of 76.7% of the cases were correctly classified when considering the entire site sample simultaneously. We also used the *leaving one out method* in which the discriminant function is first derived from only n -1 sample members, and then used to classify the observation left out. The procedure is repeated n times, each time omitting a different observation. Using this method, 72.9% of the cases were correctly classified. However, this seemingly robust performance must

Classification Results(b,c)

		NewDependent	Predicted Group Membership 1	Predicted Group Membership 2	Total
Original	Count	1	100	4	104
		2	27	2	29
	%	1	96.2	3.8	100.0
		2	93.1	6.9	100.0
Cross-validated(a)	Count	1	96	8	104
		2	28	1	29
	%	1	92.3	7.7	100.0
		2	96.6	3.4	100.0

a Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
b 76.7% of original grouped cases correctly classified.
c 72.9% of cross-validated grouped cases correctly classified.

be viewed with caution. The algorithm does a great job of identifying sites with lower rates of erosion (Damp's categories 1, 2 and 3), but a very poor job of identifying sites with high erosion rates (Damp's 4 and 5). And the latter is, after all, what we want to know.

Conclusions

While the results of this initial statistical analysis may leave much to be desired, they clearly point to the fact that no single variable can be reliably employed to determine which sites should be "in" or "out" of consideration for mitigation. The HECRAS data will undoubtedly be useful in other analytic contexts, but they represent the least powerful of the variables here addressed in assessing affects on archeological sites.

It does seem feasible that more robust prediction can be realized with 1) additional data mining of the NPS database, 2) continued refinement of monitoring protocols and 3) the use of statistical methods with less stringent requirements (e.g., logistic regression), thereby allowing categorical, ordinal and binary data into the mix of independent variables. What must be understood is that the requirements of proposed statistical models be stated at the outset of protocol development. This is the critical, though frequently ignored, nexus between data gathering and data analysis that is of paramount importance to the construction of a viable research design.

References

Damp, Jonathan, Joel Pederson and Gary O'Brien
2007 Geoarchaeological Investigations and an Archeological Treatment Plan for 151 Sites in the Grand Canyon, Arizona. Ms. On file, Bureau of Reclamation, Upper Colorado Regional Office. Salt Lake City.

Klecka, William R.
1980 Discriminant Anaysis. Sage University Paper 19. Newbury Park

Landau, Sabine and Brian S. Everitt
2004 A Handbook of Statistical Analyses using SPSS. Chapman and Hall/CRC. Boca Raton.